

**Technical Report
1085**

Track Association with Bayesian Probability Theory

M.B. Hurley

1 October 2003

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS



Prepared for the Army under Air Force Contract F19628-00-C-0002.

Approved for public release; distribution is unlimited.

20031104 016

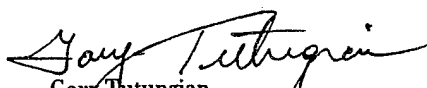
This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the Department of the Army, RTS/USAKA, under Air Force Contract F19628-00-C-0002.

This report may be reproduced to satisfy needs of U.S. Government agencies.

The ESC Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER


Gary Tutungian
Administrative Contracting Officer
Plans and Programs Directorate
Contracted Support Management

Non-Lincoln Recipients

PLEASE DO NOT RETURN

Permission is given to destroy this document
when it is no longer needed.

Massachusetts Institute of Technology
Lincoln Laboratory

Track Association with Bayesian Probability Theory

Michael B. Hurley
Group 33

Technical Report 1085

1 October 2003

Approved for public release; distribution is unlimited.

ABSTRACT

Data association and track association algorithms have been developed over the course of thirty years. Almost all technical papers that describe these association algorithms have begun the derivations by adopting log-likelihood ratios as the measure of association. At best, this starting point has obscured the assumptions necessary to use log-likelihood ratios. At worst, the log-likelihood ratios have been improperly defined. This report provides the first known derivation of a track association algorithm from the first principles of Bayesian probability theory. By starting with first principles, all the assumptions that are necessary to derive an association algorithm are explicitly stated as the derivation proceeds. The correct form for the log-likelihood ratios is obtained later in the derivation and can be traced back to first principles. The pitfalls and deficiencies of poorly performing association algorithms are identified easily by comparing the algorithms with the full derivation. These deficiencies arise from such mistakes as the incorrect definition of the log-likelihood ratio, poor selection of the probability density functions, incorrect construction of the cost matrices, and the application of an algorithm to a system that violates the assumptions that were adopted during the algorithm construction. In addition, the firm grounding in Bayesian probability theory provides the means to easily extend the derivation to produce more complex association algorithms, such as feature-aided track association algorithms. The basic derivation provided in this report makes it clear that the ensemble of track association algorithms is much more extensive than most data fusion researchers would believe. These algorithms can be created by simply changing any of the derivation assumptions or the probability density functions.

TABLE OF CONTENTS

	Page
Abstract	iii
1. INTRODUCTION	1
2. INITIAL DERIVATION	3
3. LINEAR ASSIGNMENT	11
4. PROBABILITY ESTIMATION	15
4.1 Association with Other features beyond metrics	19
4.2 Spectral feature spaces	22
5. SUMMARY	23
References	25

1. INTRODUCTION

During the critical review of a track association algorithm, a number of severe flaws were discovered in the algorithm design that would prevent the algorithm from delivering robust performance. In the course of working on track association for a fielded, real-time, data fusion system many pitfalls were encountered. At the time of the review, it was believed that the state of the art in the construction of track association algorithms had advanced to the point that the deficiencies in the reviewed algorithm would have been avoided during the design phase. While reviewing the literature on track association algorithms, it was found that the state of the art was not as advanced as had been supposed. Different approaches by different authors [1], [2], [3], [4], [5], [6], [7] were found that could not be considered a unified approach to track association, in spite of a 40-year history of development in the field [8], [9]. To resolve the disagreements between the algorithm reported here and those of various authors, the track association algorithm was rederived by starting with a basic probability equation. The inspiration for this approach was a graduate thesis [10]. This derivation shows that many of the track association algorithms are only partially correct (including that reported here) and suffer from deficiencies. In many cases, fortuitous selection of additional assumptions, such as the form of the prior probabilities, led to some algorithms performing correctly for the intended application. The derivation also gives insight into why some of the algorithms did not perform as well as expected and led to some authors choosing to adopt tuning parameters to get better performance. This new derivation lays out all the assumptions that are required to attain a straightforward track association algorithm and also provides a template for others who may wish to adopt assumptions that are different from those adopted here.

Most track association algorithms can be decomposed into a two-step algorithm. The first step is to construct a matrix of association costs between two sets of tracks. The second step usually consists of running a linear assignment algorithm on the matrix to determine the optimal track associations, in terms of the overall cost. The second step is well understood with the discovery of optimal linear assignment algorithms such as the Munkres [11], Jonker-Volgenant [12], and Jonker-Volgenant-Canstefion [1], [3]. These algorithms have, for the most part, supplanted less efficient and suboptimal algorithms such as the nearest-neighbors and the greedy. The construction of the cost matrix is less well understood and the step on which this report focuses.

This report presents the detailed derivation of a track association algorithm and clearly states all assumptions as they are made. Most previous derivations have been started with the use of likelihood ratios, obscuring most of the assumptions that are necessary for the construction of an assignment algorithm. The approach taken here is to start with a simple probability function that is dependent on the hypothetical tracks, the data, and prior information. The axioms and theorems of Bayesian probability theory [13] are then used to expand the probability function into a product of simplified probability functions. The full derivation clearly outlines how to develop variations and extensions to the standard metric track association algorithm. One variation of current interest to researchers is the extension of the track association algorithm from metrics-only to feature-aided. The simplest of these extensions are presented in this report. More advanced feature-aided track association algorithms require specific knowledge about the types of sensors and the types of targets that are involved in the association.

2. INITIAL DERIVATION

An algorithm design that is based upon a firm theoretical foundation avoids many of the problems that are almost always encountered if the design is based upon ad hoc principles. This statement is especially true for track association algorithms. The theoretical foundation provides information to the algorithm designer on how to construct track association algorithms, allows the designer to know when he is departing from theory, and allows the designer to understand why a track association algorithm might be failing.

Probability theory is the theoretical foundation for the track association algorithm derived here. The starting point could be called the probability of everything, given as

$$P(H, D, I) \quad , \quad (1)$$

where H is a hypothesis, D are the data, and I is the prior information. The probability of the prior information can be expanded into two probabilities,

$$P(H, D, I) = P(H, D | I)P(I) \quad . \quad (2)$$

where one is a conditional probability, dependent upon the parameter of the unconditional probability. This expansion is based on a standard axiom of probability theory.

One can work toward Bayes' rule by generating the equality,

$$P(H | D, I)P(D | I)P(I) = P(D | H, I)P(H | I)P(I) \quad , \quad (3)$$

with additional expansion. This relationship equates two different expansions of the conditional dependencies for the data and the hypothesis.

The identical $P(I)$ terms can be canceled from both sides of the equality, and both sides can be divided by $P(D | I)$ to arrive at a formulation of Bayes rule, where the probability of the prior information is not considered to be important:

$$P(H | D, I) = \frac{P(D | H, I)P(H | I)}{P(D | I)} \quad . \quad (4)$$

Equation (4) gives a function for the probability of the hypothesis, conditioned on the data and the prior information. The goal is to find the hypothesis that maximizes this probability, conditioned on the measured data and the prior information. The denominator on the right is usually calculated with the equation

$$P(D | I) = \sum_H P(D | H, I) P(H | I) \quad , \quad (5)$$

although the term can be ignored for most hypothesis ensembles because it is often a normalization term.

The specific problem of interest is to determine the best association between the sets of tracks from two different sensors. Association is not considered between more than two sensors for this derivation. The track sets are defined as D_1 and D_2 . The basic equation expands to

$$P(H | D_1, D_2, I) = \frac{P(D_1, D_2 | H, I) P(H | I)}{P(D_1, D_2 | I)} \quad . \quad (6)$$

A complication with track association is that one sensor may track an object that the other sensor does not detect. Missing data need to be represented in our probability equation, just like measured data are represented. We will define missing data as \bar{D}_s , where the subscript s denotes either sensor one or sensor two. The parameters expand further to

$$P(H | D_1, \bar{D}_1, D_2, \bar{D}_2, I) = \frac{P(D_1, \bar{D}_1, D_2, \bar{D}_2 | H, I) P(H | I)}{P(D_1, \bar{D}_1, D_2, \bar{D}_2 | I)} \quad . \quad (7)$$

The equation can continue to be expanded to additional conditional dependencies,

$$P(H | D, I) = \frac{P(\bar{D}_1, \bar{D}_2 | D_1, D_2, H, I) P(D_1, D_2 | H, I) P(H | I)}{P(D_1, \bar{D}_1, D_2, \bar{D}_2 | I)} \quad , \quad (8)$$

and to

$$P(H | D, I) = \frac{P(\bar{D}_1, \bar{D}_2 | D_1, D_2, H, I) P(D_1 | D_2, H, I) P(D_2 | H, I) P(H | I)}{P(D_1, \bar{D}_1, D_2, \bar{D}_2 | I)} \quad . \quad (9)$$

The term on the left has been simplified to keep the equation to one line.

The first assumption made is that

(A1) the tracks detected by one sensor are statistically independent of the tracks detected by the other sensor.

The equation simplifies slightly to

$$P(H | D, I) = \frac{P(\bar{D}_1, \bar{D}_2 | D_1, D_2, H, I) P(D_1 | H, I) P(D_2 | H, I) P(H | I)}{P(D_1, \bar{D}_1, D_2, \bar{D}_2 | I)} \quad . \quad (10)$$

The missing data terms are expanded to additional conditional dependencies:

$$P(H | D, I) = \frac{P(\bar{D}_1 | D_1, D_2, \bar{D}_2, H, I)P(\bar{D}_2 | D_1, D_2, H, I)P(D_1 | H, I)P(D_2 | H, I)P(H | I)}{P(D_1, \bar{D}_1, D_2, \bar{D}_2 | I)} \quad (11)$$

The second assumption made is that

(A2) the tracks missed by one sensor are statistically independent of the tracks missed or detected by the other sensor,

which leads to

$$P(H | D, I) = \frac{P(\bar{D}_1 | D_1, H, I)P(\bar{D}_2 | D_2, H, I)P(D_1 | H, I)P(D_2 | H, I)P(H | I)}{P(D_1, \bar{D}_1, D_2, \bar{D}_2 | I)} \quad (12)$$

A reason to keep the missing data conditionally dependent on the measured data of the same sensor is that it can account for the case when there are multiple, closely spaced, hypothetical objects, and the sensor measures the multiple, closely spaced objects as one detected object because of the sensor resolution. An approximation for target association might consider this case to be equivalent to the case when only one hypothetical object was detected and the other hypothetical objects were not. A different approach to handling hypothetical, closely spaced objects is to allow for the possibility that a detected track may represent multiple hypothetical objects. The $P(D | H, I)$ terms will have to be constructed to account for the possibility that a single detected object might be multiple, unresolved, hypothetical objects. Instead of worrying about the additional complexity that unresolved hypothetical objects add to the derivation of a track association algorithm, it is assumed that

(A3) missing data are not conditionally dependent on the measured data for a sensor.

The equation simplifies to

$$P(H | D, I) = \frac{P(\bar{D}_1 | H, I)P(\bar{D}_2 | H, I)P(D_1 | H, I)P(D_2 | H, I)P(H | I)}{P(D_1, \bar{D}_1, D_2, \bar{D}_2 | I)} \quad (13)$$

Though far from obtaining a track association algorithm, one can account for three assumptions that were made about the nature of the sensor data and the hypothetical objects. More assumptions will be made before an association algorithm is finally produced.

Because the problem considered here involves associating multiple targets from two sensors, details are added to the data and the possible hypotheses. The sensor data and the hypothesis can be considered to consist of a set of objects:

$$D_1 = \{d_{11}(x), d_{12}(x), d_{13}(x), \dots, d_{1I}(x)\} \quad , \quad (14)$$

$$D_2 = \{d_{21}(x), d_{22}(x), d_{23}(x), \dots, d_{2J}(x)\} \quad , \quad (15)$$

$$H = \{h_1(x), h_2(x), h_3(x), \dots, h_K(x), n(x)\}, a(\{i, \bar{D}_1\}, \{j, \bar{D}_2\}, \{k, n\}) \quad , \quad (16)$$

with I objects detected by sensor one, J objects detected by sensor two, and K hypothetical objects. A hypothetical noise variable $n(x)$ is defined momentarily, to allow for the possibility that some detected objects may really be due to noise. The hypothesis also includes a function $a(\{i, \bar{D}_1\}, \{j, \bar{D}_2\}, \{k, n\})$ that represents the association between the detected and missed objects in the two sensor sets, the hypothetical objects, and the noise. The variable x represents the global feature space in which the sensor measurements are made. Not all data and noise functions necessarily depend upon all the dimensions of the global feature space.

To progress further toward a track association algorithm, additional statistical independence (SI) and conditional dependence (CD) assumptions are made:

- (A4) the $d_{1i}(x)$ data are SI of the $d_{2j}(x)$ data,
- (A5) the $d_{1i}(x)$ objects are SI of each other,
- (A6) the $d_{2j}(x)$ objects are SI of each other,
- (A7) the $h_h(x)$ objects are SI of each other,
- (A8) any one $d_{si}(x)$ for any given sensor is CD on only one $h_h(x)$,
- (A9) no $d_{si}(x)$ are associated with noise $n(x)$, and
- (A10) all hypothetical objects $h_h(x)$ are detected by at least one sensor.

Assumption (A9) means that the hypothetical noise element can be neglected. Assumption (A10) prohibits hypothesizing the existence of objects that are not detected by any of the sensors. There are cases where hypothesizing the existence of undetected objects might be sensible because the prior information might support the existence of these objects. This current derivation does not consider this additional complication.

Assumptions (A9) and (A10) impose a constraint on the number of hypothetical objects,

$$\max(I, J) \leq K \leq I + J \quad . \quad (17)$$

Every detected object must match a hypothetical object. A hypothetical object is associated with either a detected object from only one given sensor or detected objects from both sensors. The current task is to

determine the most probable association between the hypothetical objects and the detected objects from the two sensors, which indirectly provides for the association between the two sets of detected objects.

The denominator is neglected in the reported probability function; proportionality between hypotheses is relied upon to search for the most probable hypothesis,

$$P(H | D, I) \propto P(\bar{D}_1 | H, I)P(\bar{D}_2 | H, I)P(D_1 | H, I)P(D_2 | H, I)P(H | I) \quad (18)$$

This simplification is reasonable because the denominator is independent of the hypotheses under consideration for the association problem.

The next step is to expand the conditional probabilities for the individual objects and impose assumptions (A4) through (A10). The resulting equation contains probabilities for associations that can be grouped into three classes: hypothetical objects detected by both sensors ($q : 1, 2$), hypothetical objects detected by sensor one but not by sensor two ($r : 1, \bar{2}$), and hypothetical objects detected by sensor two but not by sensor one ($s : \bar{1}, 2$). Each hypothetical track k is assigned to an element, q , r , or s . Each sensor track $d_{1i}(x)$ or $d_{2j}(x)$ is assigned to an element, q , r , or s , as well:

$$\begin{aligned} P(H | D, I) &\propto P(H | I) \prod_{s:1,2} P(\bar{d}_1 | h_s, I) P(d_{2j(s)} | h_s, I) \\ &\quad \times \prod_{r:1,2} P(d_{1i(r)} | h_r, I) P(\bar{d}_2 | h_r, I) \\ &\quad \times \prod_{q:1,2} P(d_{1i(q)} | h_q, I) P(d_{2j(q)} | h_q, I) \end{aligned} \quad (19)$$

Next, the hypothetical objects are assumed to have definite values in the feature space x . The definite values of the hypothetical objects in the feature space are represented by μ_q , μ_r , and μ_s for K hypothetical objects.

$$\begin{aligned} P(H | D, I) &\propto P(H | I) \prod_s P(\bar{d}_1 | h_s(\mu_s), I) P(d_{2j(s)} | h_s(\mu_s), I) \\ &\quad \times \prod_r P(d_{1i(r)} | h_r(\mu_r), I) P(\bar{d}_2 | h_r(\mu_r), I) \\ &\quad \times \prod_q P(d_{1i(q)} | h_q(\mu_q), I) P(d_{2j(q)} | h_q(\mu_q), I) \end{aligned} \quad (20)$$

The hypothesis information consists of the associations between the hypothetical objects and the sensor objects, and the hypothetical object locations, μ_q , μ_r , and μ_s in the feature space. With the probability theory approach, the locations are considered to be nuisance parameters and not necessary information if interest is only in the optimal assignment. The standard approach is to sum or integrate out the nuisance parameters, which here is summation or integration over all possible values (hypotheses) of μ_q , μ_r , and μ_s locations. Summation is used, with the viewpoint that the feature space is finite and

numerable. Continuous feature spaces are discussed later because of additional, complicating considerations. The summation is mathematically represented as

$$P(h(a_K) | D, I) = \sum_{\mu_1} \cdots \sum_{\mu_K} P(H(a_K, \mu_1, \dots, \mu_K) | D, I) \quad , \quad (21)$$

where a_K represents one association list between the hypothetical objects and sensor objects. Additional subscripts or parameters are left off for now to avoid complicating the equations. The association, a_K , can be considered to be an element in the set of all possible associations, A_K . To reiterate, each summation is over all the values that μ_q , μ_r , and μ_s can assume. The probability function expands to

$$\begin{aligned} P(h(a_K) | D, I) &\propto \sum_{\mu_1} \cdots \sum_{\mu_K} P(H(a_K, \mu_1, \dots, \mu_K) | I) \prod_s P(\bar{d}_1 | h_s(\mu_s), I) P(d_{2j(s)} | h_s(\mu_s), I) \\ &\quad \times \prod_r P(d_{1i(r)} | h_r(\mu_r), I) P(\bar{d}_2 | h_r(\mu_r), I) \\ &\quad \times \prod_q P(d_{1i(q)} | h_q(\mu_q), I) P(d_{2j(q)} | h_q(\mu_q), I) \end{aligned} \quad , \quad (22)$$

Note that the probability of the hypothesis, conditioned on the prior data, may be a function of the summands and can affect the resulting probability. The next assumption made is that

(A11) the hypothesis parameters in $P(H(a_K, \mu_1, \dots, \mu_K) | I)$ are statistically independent of one another and are thus separable into the product, $P(h(a_K) | I) P(h(\mu_1) | I) \dots P(h(\mu_K) | I)$,

leading to another simplification,

$$\begin{aligned} P(h(a_K) | D, I) &\propto \sum_{\mu_1} \cdots \sum_{\mu_K} P(h(a_K) | I) \prod_s P(h(\mu_s) | I) P(\bar{d}_1 | h_s(\mu_s), I) P(d_{2j(s)} | h_s(\mu_s), I) \\ &\quad \times \prod_r P(h(\mu_r) | I) P(d_{1i(r)} | h_r(\mu_r), I) P(\bar{d}_2 | h_r(\mu_r), I) \\ &\quad \times \prod_q P(h(\mu_q) | I) P(d_{1i(q)} | h_q(\mu_q), I) P(d_{2j(q)} | h_q(\mu_q), I) \end{aligned} \quad . \quad (23)$$

Note that a given hypothetical location really only matches with one of the probability duos inside one of the three product terms. The summations over the locations can be moved inside the products,

$$\begin{aligned}
P(h(a_K) | D, I) &\propto P(h(a_K) | I) \prod_s \sum_{\mu_s} P(h(\mu_s) | I) P(\bar{d}_1 | h_s(\mu_s), I) P(d_{2j(s)} | h_s(\mu_s), I) \\
&\times \prod_r \sum_{\mu_r} P(h(\mu_r) | I) P(d_{1i(r)} | h_r(\mu_r), I) P(\bar{d}_2 | h_r(\mu_r), I) \\
&\times \prod_q \sum_{\mu_q} P(h(\mu_q) | I) P(d_{1i(q)} | h_q(\mu_q), I) P(d_{2j(q)} | h_q(\mu_q), I)
\end{aligned} \tag{24}$$

The use of the characters q , r , and s for three different products obscures this relationship.

The next assumption made is that

(A12) the assignment hypotheses $h(a_K)$ are conditionally independent of the prior information, I , and the probability $P(h(a_K) | I)$ is uniformly distributed across the assignment hypotheses.

This term can be ignored, and leads to

$$\begin{aligned}
P(h(a_K) | D, I) &\propto \prod_s \sum_{\mu_s} P(h(\mu_s) | I) P(\bar{d}_1 | h_s(\mu_s), I) P(d_{2j(s)} | h_s(\mu_s), I) \\
&\times \prod_r \sum_{\mu_r} P(h(\mu_r) | I) P(d_{1i(r)} | h_r(\mu_r), I) P(\bar{d}_2 | h_r(\mu_r), I) \\
&\times \prod_q \sum_{\mu_q} P(h(\mu_q) | I) P(d_{1i(q)} | h_q(\mu_q), I) P(d_{2j(q)} | h_q(\mu_q), I)
\end{aligned} \tag{25}$$

This equation has the form that will be used to estimate the most probable track assignment.

3. LINEAR ASSIGNMENT

Equation (25) can be used to make the intended connection to linear programming and linear assignment algorithms. If the negative log of Equation (25) is taken, the products of probabilities convert to summations of log-probabilities.

$$\begin{aligned} \ln P(h(a_K) | D, I) \propto & \sum_s \ln \sum_{\mu_s} P(h(\mu_s) | I) P(\bar{d}_1 | h_s(\mu_s), I) P(d_{2j(s)} | h_s(\mu_s), I) \\ & + \sum_r \ln \sum_{\mu_r} P(h(\mu_r) | I) P(d_{1i(r)} | h_r(\mu_r), I) P(\bar{d}_2 | h_r(\mu_r), I) \quad (26) \\ & + \sum_q \ln \sum_{\mu_q} P(h(\mu_q) | I) P(d_{1i(q)} | h_q(\mu_q), I) P(d_{2j(q)} | h_q(\mu_q), I) \end{aligned}$$

With the equation transformed to a summation, the problem of finding the assignment list can be accomplished with linear programming algorithms [2, 3, 11, 12]. This procedure avoids the need to calculate all the permutations of assignments to find the most probable. Instead, a matrix can be constructed and a linear assignment algorithm can be used to determine the most probable assignment a_K without having to evaluate every permutation of the probability products.

Shorter representations of the negative logarithms of the summations will prove useful in the continued derivation:

$$\ell_{1j} = -\ln \sum_{\mu_s} P(h(\mu_s) | I) P(\bar{d}_1 | h_s(\mu_s), I) P(d_{2j(s)} | h_s(\mu_s), I) \quad (27)$$

$$\ell_{i2} = -\ln \sum_{\mu_r} P(h(\mu_r) | I) P(d_{1i(r)} | h_r(\mu_r), I) P(\bar{d}_2 | h_r(\mu_r), I) \quad (28)$$

$$\ell_{ij} = -\ln \sum_{\mu_q} P(h(\mu_q) | I) P(d_{1i(q)} | h_q(\mu_q), I) P(d_{2j(q)} | h_q(\mu_q), I) \quad (29)$$

A linear assignment matrix can be constructed for Equation (26):

$$M = \begin{pmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1J} & \ell_{1\bar{2}} & \infty & \cdots & \infty \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2J} & \infty & \ell_{2\bar{2}} & \cdots & \infty \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \ell_{I1} & \ell_{I2} & \cdots & \ell_{IJ} & \infty & \infty & \cdots & \ell_{I\bar{2}} \\ \ell_{\bar{1}1} & \infty & \cdots & \infty & 0 & 0 & \cdots & 0 \\ \infty & \ell_{\bar{1}2} & \cdots & \infty & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \infty & \infty & \cdots & \ell_{\bar{1}J} & 0 & 0 & \cdots & 0 \end{pmatrix} \quad (30)$$

This matrix has four quadrants. The upper-left quadrant contains the detected track association costs. The lower-left and upper-right quadrants contain the missed detection costs on the diagonals. The off-diagonal elements of these submatrices are set to infinity to prohibit their selection as an assignment and leads to improved speed from most assignment algorithms. These two quadrants are always square matrices. The lower-right quadrant is filled with zeros to counterbalance the assignment costs in the upper-left quadrant.

An interesting aspect to the linear assignment problem is that the addition of a constant value to the elements of any row or column does not change the optimal assignment solution. This fact provides for a way to slightly reduce the complexity of the matrix. It should be noted that if different assumptions are adopted for the construction of the probability equation than have been adopted to reach this point, this simplification might not be possible. This simplification is not necessary for the linear assignment algorithm to operate but provides a way to sometimes simplify the implementation of the algorithm.

First, the thresholds on the diagonals are subtracted from the appropriate rows and columns to produce the assignment matrix,

$$M = \begin{pmatrix} \ell_{11} - \ell_{\bar{1}1} - \ell_{1\bar{2}} & \ell_{12} - \ell_{\bar{1}2} - \ell_{1\bar{2}} & \cdots & \ell_{1J} - \ell_{\bar{1}J} - \ell_{1\bar{2}} & 0 & \infty & \cdots & \infty \\ \ell_{21} - \ell_{\bar{1}1} - \ell_{2\bar{2}} & \ell_{22} - \ell_{\bar{1}2} - \ell_{2\bar{2}} & \cdots & \ell_{2J} - \ell_{\bar{1}J} - \ell_{2\bar{2}} & \infty & 0 & \cdots & \infty \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \ell_{I1} - \ell_{\bar{1}1} - \ell_{I\bar{2}} & \ell_{I2} - \ell_{\bar{1}2} - \ell_{I\bar{2}} & \cdots & \ell_{IJ} - \ell_{\bar{1}J} - \ell_{I\bar{2}} & \infty & \infty & \cdots & 0 \\ 0 & \infty & \cdots & \infty & 0 & 0 & \cdots & 0 \\ \infty & 0 & \cdots & \infty & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \infty & \infty & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \quad (31)$$

The slight drawback with this matrix is that the optimal assignment cost that is calculated by most linear assignment algorithms is not proportional to the negative log of the assignment probability unless the diagonal probabilities are retained and added back into the optimal assignment cost.

Because three of the quadrants now contain zeros or infinities, the linear assignment problem can be solved with the smaller matrix

$$M = \begin{pmatrix} \ell_{11} - \ell_{\bar{1}1} - \ell_{1\bar{2}} & \ell_{12} - \ell_{\bar{1}2} - \ell_{1\bar{2}} & \cdots & \ell_{1J} - \ell_{\bar{1}J} - \ell_{1\bar{2}} \\ \ell_{21} - \ell_{\bar{1}1} - \ell_{2\bar{2}} & \ell_{22} - \ell_{\bar{1}2} - \ell_{2\bar{2}} & \cdots & \ell_{2J} - \ell_{\bar{1}J} - \ell_{2\bar{2}} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{I1} - \ell_{\bar{1}1} - \ell_{I\bar{2}} & \ell_{I2} - \ell_{\bar{1}2} - \ell_{I\bar{2}} & \cdots & \ell_{IJ} - \ell_{\bar{1}J} - \ell_{I\bar{2}} \end{pmatrix}, \quad (32)$$

and the matched-pair indices of the optimal assignment that are less than zero can be accepted as the assignments that should be made. Matched-pair indices with values greater than zero would have matched with the missed-object elements in the larger matrix.

In general, this smaller matrix is not square, so the linear assignment algorithm has to be able to find solutions to nonsquare matrices, or zero-element rows and columns have to be added to make the matrix square for linear assignment algorithms that can only solve square matrices. If the case is the latter, then matched-pair indices that end up in the added rows or columns are considered to be one object that was detected by one sensor and not detected by the other.

If the diagonal threshold pair sums, $\ell_{\bar{1}j} + \ell_{i\bar{2}}$, are all equal or chosen to be equal, the subtraction of the diagonal terms can be ignored and the upper left quadrant of Equation (30) solved with a linear assignment algorithm capable of solving rectangular matrices. If the linear assignment algorithm can solve only square matrices, additional rows or columns with values equal to the threshold pair sum are added to square the matrix to allow for a solution. Pair assignments with a cost greater than or equal to the threshold sum are matches that are rejected because the match with missed-object elements is equally or more probable.

4. PROBABILITY ESTIMATION

The basis for the association algorithm now exists; functions must be obtained for the conditional probabilities

$$P(h(\mu_k) | I) \quad , \quad (33)$$

$$P(\bar{d}_1 | h_s(\mu_s), I) \quad , \quad (34)$$

$$P(\bar{d}_2 | h_r(\mu_r), I) \quad , \quad (35)$$

$$P(d_{1i(k)} | h_k(\mu_k), I) \quad , \text{ and} \quad (36)$$

$$P(d_{2j(k)} | h_k(\mu_k), I) \quad . \quad (37)$$

Note that Equations (36) and (37) are probabilities of the data, conditioned on the hypothesis and the prior information, and not probabilities of the hypothesis, conditioned on the data and prior information. The two conditional probabilities are different. The natural inclination is to think that a multitarget tracker at a sensor would be designed to estimate the most-probable hypothetical tracks that have been conditioned on the data that the sensor has collected. For track association between sensors, the desired information is a function that gives the probability of the data, conditioned on the hypotheses. The distinction is subtle, but important. The importance depends upon the nature of the probability distribution functions that are used. The special symmetry property of the Gaussian functions allows for the distinction usually to be ignored without major consequences. Interchange of the data parameters and hypothesis parameters in the Gaussian function results in the same function. In addition, the integral of the Gaussian function over the hypothesis parameters and the integral of the Gaussian function over the data parameters are equal to one.

The first probability function selected is the probability of the hypothesis states $P(H(\mu_k) | I)$ on the feature space. Prior information may indicate that the hypothetical objects are less or more likely to occur in different regions of the feature space. Commonly, the adopted assumption is that

(A13) the prior information does not provide any indication that objects are more or less likely to occur in any given region of feature space. In other words, the probability distribution for the object locations is a uniform distribution across the metric space. This probability then can be set to either

$$P(h(\mu_k) | I) = \frac{1}{N} \quad , \quad (38)$$

for N elements of a countable feature space, or

$$\rho(h(\mu_k) | I) = \frac{1}{V} \quad , \quad (39)$$

for a volume V in a bounded, continuous feature space. For now, if the feature space is unbounded and continuous, a suitable boundary is chosen for the integrals so that the bounded, continuous equality can be used.

Functions for the remaining probabilities require additional assumptions about feature spaces, so specific cases of the association problem are examined. Real, bounded, feature spaces are reviewed first because the association problem is most often formulated with Gaussian probability density functions in a real feature space. Feature space is considered to be of an integer number of dimensions. It is assumed that the probabilities $P(d_{1i(k)} | h_k(\mu_k), I)$ and $P(d_{2j(k)} | h_k(\mu_k), I)$ are Gaussian functions. The trackers at the sensors are assumed to generate a state that provides a mean position $\langle x \rangle$ and covariance Σ . Because Gaussian functions are symmetric for interchange in μ_k and $\langle x \rangle$, the Gaussian function from the sensors can be used for the detected object probabilities,

$$G(\mu_k, \langle x \rangle, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mu_k - \langle x \rangle)^T \Sigma^{-1}(\mu_k - \langle x \rangle)\right) \quad . \quad (40)$$

Since the original calculations were carried out with numerable sets, we need to convert the summations to integrations. First we'll estimate the probability for the hypothesis location being within a small volume element Δv . With the uniform hypothesis prior assumption, the probability becomes

$$P(h(\mu_q) | I) P(d_{1i(q)} | h_q(\mu_q), I) P(d_{2j(q)} | h_q(\mu_q), I) = \frac{\Delta v_\mu}{V} G(\mu_q, \langle x_{1i} \rangle, \Sigma_{1i}) \Delta v_{1i} G(\mu_q, \langle x_{2j} \rangle, \Sigma_{2j}) \Delta v_{2j} \quad . \quad (41)$$

Note the three different Δv terms. It is important to distinguish between them because many probability calculations have gone awry when the terms have been confused. There may be only one feature space, but there are three parameter spaces. The parameter space volume elements are those parameters in the probability term that appear to the left of the '|' character.

The Δv_μ term is used to convert from summation to integration over the parameter space for μ_k . The integration of the preceding probability results in the approximation

$$P(\cdot) \approx \frac{\Delta v_{1i} \Delta v_{2j}}{V \sqrt{2\pi(\Sigma_{1i} + \Sigma_{2j})}} \exp\left(-\frac{1}{2}(\langle x_{1i} \rangle - \langle x_{2j} \rangle)^T (\Sigma_{1i} + \Sigma_{2j})^{-1} (\langle x_{1i} \rangle - \langle x_{2j} \rangle)\right) \quad (42)$$

when V is sufficiently large. The negative log of this term is

$$\ell_{ij} = \frac{1}{2} (\langle x_{1i} \rangle - \langle x_{2j} \rangle)^T (\Sigma_{1i} + \Sigma_{2j})^{-1} (\langle x_{1i} \rangle - \langle x_{2j} \rangle) + \frac{1}{2} \ln(2\pi(\Sigma_{1i} + \Sigma_{2j})) - \ln(\Delta v_{1i} \Delta v_{2j} / V). \quad (43)$$

The first term of Equation (43) is one-half the Mahalanobis distance and is sometimes called either the covariance-weighted distance between the two means or the chi-squared distance. The second term in the log-likelihood involves the combined covariances. The third term is an incremental length. The units associated with the second and third terms cancel.

Many assignment algorithms have been constructed that use only the Mahalanobis distance term, leading to algorithms that perform poorly when the covariances differ from track to track, especially within the set of tracks for one sensor. When this term is neglected, the less accurate tracks steal associations from the more accurate tracks because the Mahalanobis distance is not directly proportional to the association probability. Mathematically, the Mahalanobis distance is only appropriate when the second term of Equation (43) is constant for all combinations of i and j . Appropriate simple changes to the threshold log-probabilities l_{1j} and l_{i2} have to be made if the Mahalanobis distance is used instead of a probabilistic cost. Generally, the threshold is selected as a limit on the number of standard deviations before track associations are unacceptable.

A simple, one-dimensional numerical example can illustrate the problem with the inappropriate use of the Mahalanobis distance. Assume that sensor one reports one object with mean and covariance (0,1). Sensor two reports two objects with means and covariances (1,1) and (7,100); the second object's error is ten times larger than the first track's error. This difference in accuracy is not uncommon for sensor track data. The Mahalanobis distances for the two possible associations are 0.2500 and 0.2426. The Mahalanobis distance would select the less accurate track from sensor two as the one that associates with the single track from sensor one. The two association probabilities are 0.2196 and 0.0311. The odds are 7:1 that the more accurate track is the one that associates with sensor one's track, but the use of only the Mahalanobis distance selects the less accurate track. Constructing an assignment algorithm that uses only Mahalanobis distance will generate improbable assignments.

The threshold terms, l_{1j} and l_{i2} , must be determined, with somewhat more difficulty than that of selecting thresholds for Mahalanobis-distance assignment algorithms. For these algorithms, designers usually select the number of standard deviations to use to set the threshold level. Threshold selection is more difficult with the correct formulation with log-probabilities, but provides a more powerful technique for setting thresholds because additional factors that influence detection can be incorporated into the threshold function.

If prior information is available on the sensor's sensitivity to detecting objects, this sensitivity can be taken into account in the construction of the functions for $P(\bar{d}_1 | h_s(\mu_s), I)$ and $P(\bar{d}_2 | h_r(\mu_r), I)$. This sensitivity can provide very powerful information with significant impact on the optimal assignment. In this example, it is assumed that this information is not available, and suitable approximations to the threshold terms are constructed. The detected-object probabilities are assumed to be Gaussian functions

$$P(h(\mu_s) | I)P(\bar{d}_1 | h_s(\mu_s), I)P(d_{2j(s)} | h_s(\mu_s), I) = \frac{\Delta v_\mu}{V} P(\bar{d}_1 | h_s(\mu_s), I)G(\mu_s, \langle x_{2j} \rangle, \Sigma_{2j})\Delta v_{2j} \quad (44)$$

The probability of a missed detection is assumed to be independent of the hypothetical location. The integration over Δv_μ , where V is sufficiently large, leads to the result

$$\ell_{\bar{1}j} = -\ln(P(\bar{d}_1)) - \ln(\Delta v_{2j}/V) \quad (45)$$

The other threshold term produces

$$P(h(\mu_r) | I)P(d_{1i(r)} | h_r(\mu_r), I)P(\bar{d}_2 | h_r(\mu_r), I) = \frac{1}{V} \Delta v_\mu G(\mu_r, \langle x_{1i} \rangle, \Sigma_{1i})P(\bar{d}_2 | h_r(\mu_r), I)\Delta v_{1i} \quad (46)$$

and

$$\ell_{\bar{1}j} = -\ln(P(\bar{d}_2)) - \ln(\Delta v_{1i}/V) \quad (47)$$

Because pairs of thresholds are compared against associations, the Δv_{1i} and Δv_{2j} terms appear once in all hypothetical association probabilities. They can be ignored. The same is not true for the volume term V from the uniform prior of the hypothesis. The threshold pairs contribute two $1/V$ terms, while the association term contributes one.

One possibility is to remove the $1/V$ term from the association log likelihood, and distribute the remaining $1/V$ term across the thresholds if the reduced linear assignment matrix is to be used:

$$\ell_{ij} = \frac{1}{2} (\langle x_{1i} \rangle - \langle x_{2j} \rangle)^T (\Sigma_{1i} + \Sigma_{2j})^{-1} (\langle x_{1i} \rangle - \langle x_{2j} \rangle) + \frac{1}{2} \ln(2\pi(\Sigma_{1i} + \Sigma_{2j})) \quad (48)$$

$$\ell_{\bar{1}j} = -\ln(P(\bar{d}_1)) + \frac{1}{2} \ln(V) \quad (49)$$

$$\ell_{\bar{1}j} = -\ln(P(\bar{d}_2)) + \frac{1}{2} \ln(V) \quad (50)$$

The volume element is problematic, especially when the derivations are extended to an unbounded feature space. All the log-likelihood functions have a limit on infinity for infinite feature spaces if Equations (43), (45), and (47) are used. If Equations (48) through (50) are used, the limit leads to the maximum number of associations being made between the tracks from the two sensors because the limits for the thresholds are infinities. One approach to this problem is to use the argument that follows.

An infinite number of hypothetical objects are assumed to be in infinite space; their density is $1/V_p$. It is next assumed that the volume associated with the density is sparse enough that the tracks observed by the sensors are individual, distinguishable objects. The probability density for a hypothetical object that is detected by a sensor is uniform over this volume. The density function integrals with this prior probability function are assumed to be reasonably approximated by the integral of two Gaussian functions over an infinite volume, with the probability densities of Gaussian functions falling off rapidly enough that the regions with zero prior probability contribute little to the integral. This process matches in spirit the derivations of Stone et al. [4], who chose that the value of V_p is a volume equivalent to three standard deviations of the combined Gaussian covariances. For example,

$$V_p \approx \left| 9 \left(\max_i (\Sigma_{1i}) + \max_j (\Sigma_{2j}) \right) \right|^{1/2} \quad (51)$$

could be an equivalent technique to estimate a region for a hypothetical object. The thresholds and association probabilities then can be calculated with V_p , replacing V in the appropriate equations.

4.1 ASSOCIATION WITH OTHER FEATURES BEYOND METRICS

The use of metric features for the association is limited by the resolution of the sensors. If metric accuracy cannot be improved, improved performance from the association algorithm can be achieved only by considering other feature information in addition to the metric. The additional measurements, or features, can be used to strengthen the association between tracks or to prevent association when it is likely that two different objects are actually very close together that should not be associated. This situation is probable to occur for closely spaced targets, one object of a type that can be observed by one sensor but not the other sensor, and the other object of a type observed by the other sensor but not the first. If there are measurable features that can provide enough probabilistic evidence that the two tracks are two different objects, then the additional features can prevent the association.

Adding feature information is relatively straightforward, at least in terms of the association algorithm. The real work lies in estimating the probabilities for the new feature information. The ease of adding association evidence arises from the nature of the logarithmic function,

$$\ln(PQ) = \ln(P) + \ln(Q) \quad (52)$$

If the probability distributions are separable,

$$P(x, f | H, I) = P(x | H, I)P(f | x, H, I) = P(x | H, I)P(f | H, I) \quad , \quad (53)$$

then the association cost matrices can be calculated independently and simply added together for the linear assignment operation. Unless there are doubts on the independence of the probability density

functions or the belief that the probabilities of the features or the metrics are not realistic, there is no multiplicative scale factor for either the metric or the feature cost matrix.

An example of feature space separability can be used to demonstrate the separation by making two assumptions:

(B1) the two feature spaces are statistically independent, and

(B2) the probability of not detecting a track is independent of the new feature space, given by f .

Assumption (B2) is made here only to provide a specific example assumes that the feature space f has no influence on the detectability of the targets. There are strong reasons not to assume (B2), if possible. It is preferable to be able to account for differences between the types of targets that the two sensors are able to detect. The added information improves the probabilities that the right assignments will be selected. The potential difficulty is that the missed-detection probability may not be separable into two products, complicating the construction of the total cost matrix. If the probabilities are fully separable for all terms, then the total cost matrix can be constructed with the addition of independent matrices, traditionally, one for the metric and one for the additional feature costs.

Given the adoption of Assumptions (B1) and (B2), the log-likelihood functions expand to

$$\ell_{\bar{1}j} = -\ln \sum_{\mu_s, f_s} P(h(\mu_s) | I) P(h(f_s) | I) P(\bar{d}_1 | h_s(\mu_s), I) P(d_{2j(s)} | h_s(\mu_s), I) P(d_{2j(s)} | h_s(f_s), I) \quad (54)$$

$$\ell_{i\bar{2}} = -\ln \sum_{\mu_r, f_r} P(h(\mu_r) | I) P(h(f_r) | I) P(d_{1i(r)} | h_r(\mu_r), I) P(d_{1i(r)} | h_r(f_r), I) P(\bar{d}_2 | h_r(\mu_r), I) \quad (55)$$

$$\begin{aligned} \ell_{ij} = & -\ln \sum_{\mu_q, f_q} P(h(\mu_q) | I) P(h(f_q) | I) P(d_{1i(q)} | h_q(\mu_q), I) P(d_{1i(q)} | h_q(f_q), I) \\ & \times P(d_{2j(q)} | h_q(\mu_q), I) P(d_{2j(q)} | h_q(f_q), I) \end{aligned} \quad (56)$$

and separate to

$$\begin{aligned} \ell_{\bar{1}j} = & -\ln \sum_{\mu_s} P(h(\mu_s) | I) P(\bar{d}_1 | h_s(\mu_s), I) P(d_{2j(s)} | h_s(\mu_s), I) \\ & - \ln \sum_{f_s} P(h(f_s) | I) P(d_{2j(s)} | h_s(f_s), I) \end{aligned} \quad (57)$$

$$\begin{aligned} \ell_{i\bar{2}} = & -\ln \sum_{\mu_r} P(h(\mu_r) | I) P(d_{1i(r)} | h_r(\mu_r), I) P(\bar{d}_2 | h_r(\mu_r), I) \\ & - \ln \sum_{f_r} P(h(f_r) | I) P(d_{1i(r)} | h_r(f_r), I) \end{aligned} \quad (58)$$

$$\begin{aligned}\ell_{ij} = & -\ln \sum_{\mu_q} P(h(\mu_q) | I) P(d_{1i(q)} | h_q(\mu_q), I) P(d_{2j(q)} | h_q(\mu_q), I) \\ & -\ln \sum_{f_q} P(h(f_q) | I) P(d_{1i(q)} | h_q(f_q), I) P(d_{2j(q)} | h_q(f_q), I)\end{aligned}\quad (59)$$

If the prior probability $P(h(f_k))$ is a uniform prior of $1/F$, then the feature costs simplify to

$$\ell_{f1j} = \ell_{f2} = \ln F \quad (60)$$

$$\ell_{fij} = -\ln \left(\frac{1}{F} \sum_{f_q} P(d_{1i(q)} | h_q(f_q), I) P(d_{2j(q)} | h_q(f_q), I) \right) \quad (61)$$

The similarities of the feature probability distributions increase or decrease the likelihood of association between tracks.

Note that if all the tracks provide no evidence for or against the preference of any of the object types, the probabilities are the uniform distribution, $1/F$. The feature costs then reduce to

$$\ell_{fij} = 2 \ln(F) \quad , \quad (62)$$

which produces a matrix that provides no additional evidence for or against association. The likelihood for the lack of feature evidence in Equation (62) is balanced by the product of the two missed-detection costs in Equation (60).

This example holds true only for those cases where the probability of detection is independent of the feature subspace. If the probability of detection depends on the feature subspace, then costs are different and provide more information as to what objects should or should not be associated. It provides for the ability to prevent association between closely spaced targets if the evidence supports sufficiently different target classes, especially if the two sensors are better able to detect different classes of targets.

One difficulty with additional features is that the entire cost matrix must be populated with estimates that are determined from probability density functions. If feature information is missing from a track, a suitable probability estimate still must be selected, such as a uniform probability distribution. The threshold probabilities need to be calculated for tracks with missing feature data as well.

A weakness with using class or identification (ID) probability vectors as the additional feature subspace is that the feature space is usually not large enough to have much of an influence on the optimal associations in comparison to the influence of the metric association matrix. The metric space is usually a real space of two, three, six, or even nine dimensions. The class or ID space tends to be an integer space with a finite span, usually with only a few tens of different classes or IDs. Even if the feature space is

relatively large, the probabilities must be extremely high or low to drive the associations that are made in comparison to the metrics.

4.2 SPECTRAL FEATURE SPACES

Many researchers have attempted using spectral information from radar cross section or radiometric intensity measurements as a means to associate tracks between sensors, but in the opinion of this writer, very few efforts have come close to succeeding. Their knowledge of track association and its relationship to probabilities makes it easier to sense why this technique has been difficult. One reason for the difficulty is the need for a way to transform a pair of spectra from two tracks into an association probability:

$$P(t_{1i} \equiv t_{2j} | S_{1i} S_{2j}) \quad (63)$$

for two spectra, S_{1i} and S_{1j} . This need has generally been neglected in the problem definition. Another reason for the difficulty with spectra association is that the frequency peaks in the two spectra are only loosely correlated to each other. The observed peaks f_o are usually a function of the spin and precession frequencies,

$$f_o = Nf_s + Mf_p \quad , \quad (64)$$

where N and M are positive, 0, or negative integers, as long as f_o , f_s , and f_p are positive. The observed frequency peaks vary between sensors, depending on the nature of the target, sensor characteristics, and viewing geometries.

The desire to use spectral data for track association is partially driven by human nature. The peaks are often easy to discern. Many data in the frequency plot lead to the assumption that there are a large number of possible configurations to the plots and that very small association probabilities should be obtainable when the sensors are observing different, dissimilar objects. Reconsidering the argument that the frequency spectra contain a lot of information (data), it could be argued that the actual time-series measurements should have even more data that can be used to estimate even smaller association probabilities for different, dissimilar objects. Most researchers quickly recognize that it is difficult to estimate association probabilities with raw data because the information in the raw data plots is more difficult to discern, and that it is even more difficult to construct an association measure for the information in the measurement data from two tracks. One way to simplify the construction of a probabilistic association measure is to create algorithms that extract the rotational and precessional frequency, simplifying the generation of probability density functions for the spin and precession frequencies. The algorithm must recognize that the frequencies from one sensor may be a rational fraction of the corresponding measured frequencies observed for the track from the other sensor.

5. SUMMARY

This report has presented a new approach to deriving track association algorithms by starting with a basic probability function and using the axioms and theorems of Bayesian probability theory to expand the basic function into the necessary format for track association. The technique accounts for all the assumptions that are necessary to arrive at a derivation of a track association algorithm. The derivation carried out here results in a typical track association algorithm. It highlights the problems that are often encountered with track association algorithms. The derivation also provides a template that can be used to construct other track association algorithms; the interested algorithm developer can change the assumptions and work through a new derivation to arrive at a different algorithm. The developer can also derive more complex algorithms with the same derivation and assumptions, but can use more complex (informative) probability density functions. Variations beyond the simple examples provided here can include changing such things as the form of the track probability density functions, detection probability functions, feature spaces, and feature space decomposition.

Additional knowledge about the characteristics of the objects and the sensors is useful if it can be incorporated into the association algorithm. If accurate prior information about the objects is available, the information can be incorporated in the prior probability functions used in the derivation. The derivation has to be redone to get the appropriate functions for construction of the association matrix. If the detection characteristics of the sensors are known and can be embodied in the detection probability functions, better association performance can be achieved.

If the intended application severely violates any of the assumptions used to derive the association matrix equations, then the derivation must be redone with the new set of assumptions in order to derive a workable association algorithm. Whether the new assumptions lead to a linear programming algorithm depends on the new assumptions. A new assignment algorithm can be constructed by starting with Equation (1) and carrying out the new derivation with a different set of assumptions and prior information.

Although the focus of this report has been on the derivation of track association algorithms with Bayesian probability theory and linear assignment algorithms, there is no intention to imply that this approach is the only way to develop track association algorithms. It is certainly possible to construct track association algorithms with other theories, such as statistical decision [14] or Dempster-Shafer evidential reasoning [15]. Little work has been done with track association algorithms in these areas at this time and most track association algorithms that have been developed to date can be more easily related to the track association algorithm that has been outlined in this report. Bayesian probability theory provides the best guidance to the construction of track association algorithms.

REFERENCES

- 1 D. B. Reid, "An Algorithm for Tracking Multiple Targets," *IEEE Trans. Automat. Contr.* **AC-24** (6), December 1979
- 2 O. Drummond D.A. Castañon, and M.S. Bellowing, "Comparison of 2-D Assignment Algorithms for Rectangular, Floating Point cost Matrices," *Proc. SDI Panels on Tracking* (4), December 1990.
- 3 D. A. Castañon, "New Assignment Algorithms for Data Association," SPIE vol. 1698 *Signal and Data Processing of Small Targets*, 1992, 313–323.
- 4 S. Mori and C-Y Chong, "Effects of Unpaired Objects and Sensor biases on Track-to-Track Association: Problems and Solutions," *Proc. MSS Sensor and Data Fusion 2000*, I, CD-ROM, 2000.
- 5 S. Mori and C-Y Chong, "BMD Mid-Course Object Tracking: Track fusion under Asymmetric Conditions," *Proc. National Symposium for Sensor and Data Fusion*, I, CD-ROM, 2001.
- 6 C-Y Chong and S. Mori, "Track Fusion for NMD Mid-Course Tracking," *Proc. National Symposium for Sensor and Data Fusion*, ERIM, CD-ROM, 2002.
- 7 L. D. Stone, M. L. Williams, and T. M. Tran, "Track-to-Track Association and Bias Removal," *Signal and Data Processing of Small Targets 2002*, O. E. Drummond (ed.), *Proc. SPIE*, **4728**, 315–329.
- 8 S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, Norwood, Mass., 1999, p. 10.
- 9 R. W. Sittler, "An Optimal Data Association Problem in Surveillance Theory," *IEEE Trans. Mil. Electron.*, MIL-8, April 1964, p. 125–139.
- 10 G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*, Springer-Verlag, New York, 1988.
- 11 F. Bourgeois and J. C. Lassalle, "An Extension of the Munkres Algorithm for the Assignment Problem to Rectangular Matrices," *Comm. ACM*, **14**, December 1971, p. 802–804.
- 12 R. Jonker and A. Volgenant, "A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems," *Computing*, **38**, 1987, p. 325–340.
- 13 H. Jeffreys, *Theory of Probability*, 3rd ed, Clarendon Press, Oxford University, England 1961.

- 14 D. Middleton, *An Introduction to Statistical Communication Theory*, IEEE Press, Piscataway, N.J., 1992.
- 15 G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, N.J., 1976.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 1 October 2003	3. REPORT TYPE AND DATES COVERED Technical Report		
4. TITLE AND SUBTITLE Track Association with Bayesian Probability Theory		5. FUNDING NUMBERS C—F19628-00-C-0002		
6. AUTHOR(S) M.B. Hurley				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Lincoln Laboratory, MIT 244 Wood Street Lexington, MA 02420-9108		8. PERFORMING ORGANIZATION REPORT NUMBER TR-1085		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) RTS Support Directorate USA Space & Missile Command SMDC-AC-K-S1 P.O. Box 1500 Huntsville, AL 35807-3801		10. SPONSORING/MONITORING AGENCY REPORT NUMBER ESC-TR-2003-057		
11. SUPPLEMENTARY NOTES None				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) Data association and track association algorithms have been developed over the course of thirty years. Almost all technical papers that describe these association algorithms have begun the derivations by adopting log-likelihood ratios as the measure of association. At best, this starting point has obscured the assumptions necessary to use log-likelihood ratios. At worst, the log-likelihood ratios have been improperly defined. This report provides the first known derivation of a track association algorithm from the first principles of Bayesian probability theory. By starting with first principles, all the assumptions that are necessary to derive an association algorithm are explicitly stated as the derivation proceeds. The correct form for the log-likelihood ratios is obtained later in the derivation and can be traced back to first principles. The pitfalls and deficiencies of poorly performing association algorithms are identified easily by comparing the algorithms with the full derivation. These deficiencies arise from such mistakes as the incorrect definition of the log-likelihood ratio, poor selection of the probability density functions, incorrect construction of the cost matrices, and the application of an algorithm to a system that violates the assumptions that were adopted during the algorithm construction. In addition, the firm grounding in Bayesian probability theory provides the means to easily extend the derivation to produce more complex association algorithms, such as feature-aided track association algorithms. The basic derivation provided in this report makes it clear that the ensemble of track association algorithms is much more extensive than most data fusion researchers would believe. These algorithms can be created by simply changing any of the derivation assumptions or the probability density functions.				
14. SUBJECT TERMS			15. NUMBER OF PAGES 108	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Same as Report	